



Almost sure convergence of randomly truncated stochastic algorithms under verifiable conditions

Jérôme Lelong

► To cite this version:

Jérôme Lelong. Almost sure convergence of randomly truncated stochastic algorithms under verifiable conditions. *Statistics and Probability Letters*, 2008, 78 (16), pp.2632-2636. 10.1016/j.spl.2008.02.034 . hal-00152255

HAL Id: hal-00152255

<https://hal.science/hal-00152255>

Submitted on 6 Jun 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ALMOST SURE CONVERGENCE OF RANDOMLY TRUNCATED STOCHASTIC ALGORITHMS UNDER VERIFIABLE CONDITIONS

JÉRÔME LELONG

ABSTRACT. In this paper, we are interested in the almost sure convergence of randomly truncated stochastic algorithms. In their pioneer work, [Chen and Zhu \(1986\)](#) required that the family of the noise terms is summable to ensure the convergence. In our paper, we present a new convergence theorem which extends the already known results by making vanish this condition on the noise terms — a condition which is quite hard to check in practice. The aim of this work is to prove an almost sure convergence result of randomly truncated stochastic algorithms under easily verifiable conditions (see Theorem 1).

Key words. stochastic approximation, randomly truncated algorithms, almost sure convergence.

1. INTRODUCTION

The localisation of the zeros of a function u is a quite complicated problem for which many techniques have already been developed. The use of stochastic algorithms is widely spread for solving such problems. Stochastic algorithms are particularly well suited where some on-line parameter estimation is needed. Such algorithms go back to the pioneer work of [Robbins and Monro \(1951\)](#). They proposed to consider the following recurrence relation

$$X_{n+1} = X_n - \gamma_{n+1}u(X_n) - \gamma_{n+1}\delta M_{n+1},$$

where γ_n is a decreasing gain sequence and Δm_n the measurement error. Under certain conditions on the growth of the \mathbb{L}^2 -norm of the error, X_n converges almost surely to the unique root of u . Since their work, much attention has been drawn to the study of the theory of such recursive approximations. The first works were dealing with independent measurement error on the observations. A great effort was made in this direction to weaken the conditions imposed on both the regressive function and the noise term. Using the ordinary differential equation technique, [Kushner and Clark \(1978\)](#) proved a convergence result for a wider range of measurement noises and in particular for martingale increments.

One major drawback of these algorithms is that their convergence can only be established if the function u does not grow too quickly, namely a sub-linear behaviour is required. This is a dramatic restriction for practical applications. [Chen and Zhu \(1986\)](#) have found a way to get round the restriction by considering stochastic algorithms truncated at randomly varying bounds. Their algorithm can be written

$$X_{n+1} = X_n - \gamma_{n+1}u(X_n) - \gamma_{n+1}\delta M_{n+1} - \gamma_{n+1}p_{n+1}, \quad (1)$$

where p_n is a truncation term.

In this paper, we are concerned with the convergence of the truncated algorithm (1). Several results already exist but the hypotheses considered differ quite significantly. The first result

Date: June 6, 2007.

concerning the almost sure convergence is due to [Chen and Zhu \(1986\)](#). The convergence was also studied by [Delyon \(1996\)](#) and [Delyon et al. \(1999\)](#). The robustness of the procedure was established by [Chen et al. \(1988\)](#) under global hypotheses on the measurement error. Namely, they require that the series $\sum_n \gamma_n \delta M_n$ converges almost surely. [Delyon \(1996\)](#) has also studied the almost sure convergence under local hypotheses on the measurement noise. Here, we give a self-contained proof of the convergence under local hypotheses, that is we only assume that $\sum_n \gamma_n \delta M_n \mathbb{1}_{\{X_n \in K\}}$ converges almost surely for any compact set K . We do not impose any condition on the truncation term p_n .

First, we define the general framework and explain the algorithm developed by [Chen and Zhu \(1986\)](#). Our main result is stated in Theorem 2 in a very general way. For practical purposes, we give in Theorem 1 an easily verifiable condition under which our main result holds. This theorem is extremely valuable and dramatically extends the range of applications of randomly truncated stochastic algorithms. Finally, Section 4 is devoted to the proof of the general convergence theorem.

2. GENERAL FRAMEWORK

Let us consider a general problem consisting in finding the root of a continuous function $u: X \in \mathbb{R}^d \mapsto u(X) \in \mathbb{R}^d$, defined as an expectation on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

$$u(X) = \mathbb{E}(U(X, Z)), \quad (2)$$

where Z is a random variable in \mathbb{R}^m and U a measurable function defined on $\mathbb{R}^d \times \mathbb{R}^m$ into \mathbb{R}^d . We assume that $x \mapsto \mathbb{E}(\|U(x, Z)\|^2)$ grow faster than $\|x\|^2$, so that the convergence of the standard Robbins Monro algorithm is not guaranteed. Instead, we consider the alternative procedure proposed by [Chen and Zhu \(1986\)](#), on which we concentrate in this work.

The technique consists in forcing the algorithm to remain in an increasing sequence of compact sets. Somehow, it prevents the algorithm from blowing up during the "first" steps.

We consider an increasing sequence of compact sets $(\mathcal{K}_j)_j$

$$\bigcup_{j=0}^{\infty} \mathcal{K}_j = \mathbb{R}^d \quad \text{and} \quad \forall j, \mathcal{K}_j \subsetneq \text{int}(\mathcal{K}_{j+1}). \quad (3)$$

We also introduce $(Z_n)_n$ an independent and identically distributed sequence of random variables following the law of Z and $(\gamma_n)_n$ a decreasing sequence of positive real numbers. γ_n is often called the gain sequence. For any deterministic $X_0 \in \mathcal{K}_0$ and $\sigma_0 = 0$, we define the sequences of random variables $(X_n)_n$ and $(\sigma_n)_n$.

$$\begin{cases} X_{n+\frac{1}{2}} = X_n - \gamma_{n+1} U(X_n, Z_{n+1}), \\ \text{if } X_{n+\frac{1}{2}} \in \mathcal{K}_{\sigma_n} & X_{n+1} = X_{n+\frac{1}{2}} \quad \text{and} \quad \sigma_{n+1} = \sigma_n, \\ \text{if } X_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n} & X_{n+1} = X_0 \quad \text{and} \quad \sigma_{n+1} = \sigma_n + 1. \end{cases} \quad (4)$$

Remark 1. When $X_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n}$, one can set X_{n+1} to any measurable function of (X_0, \dots, X_n) with values in a given compact set. This existence of such a compact set is definitely essential to prove the a.s. convergence of $(X_n)_n$.

Remark 2. $X_{n+\frac{1}{2}}$ represents the iterate of the Robbins Monro algorithm at step $n+1$.

We introduce $\mathcal{F}_n = \sigma(Z_k; k \leq n)$ the σ -field generated by the random vectors Z_k , for $k \leq n$. Note that X_n is \mathcal{F}_n -measurable since X_0 is deterministic and U measurable. We can write $u(X_n) = \mathbb{E}[U(X_n, Z_{n+1}) | \mathcal{F}_n]$.

It is often more convenient to rewrite (4) as follows

$$X_{n+1} = X_n - \gamma_{n+1}u(X_n) - \gamma_{n+1}\delta M_{n+1} + \gamma_{n+1}p_{n+1} \quad (5)$$

where

$$\begin{aligned} \delta M_{n+1} &= U(X_n, Z_{n+1}) - u(X_n), \\ \text{and } p_{n+1} &= \begin{cases} u(X_n) + \delta M_{n+1} + \frac{1}{\gamma_{n+1}}(X_0 - X_n) & \text{if } X_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n}, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (6)$$

Remark 3. δM_n is a martingale increment. The case of the standard Robbins Monro algorithm corresponds to $p_n = 0$.

3. ALMOST SURE CONVERGENCE

In this section, we present a new convergence theorem that improves the result of [Chen and Zhu \(1986\)](#) who proved the almost sure convergence under global hypotheses on the series $\sum_n \gamma_{n+1}\delta M_{n+1}$ whereas we can manage the proof under local hypotheses only, namely we only assume that the function $x \mapsto \mathbb{E}(\|U(x, Z)\|^2)$ is bounded on all compact sets. Such a local hypothesis is much easier to satisfy in practical applications.

Theorem 1. *We assume that*

(A1) *There exists a unique x^* s.t. $u(x^*) = 0$ and $\forall x \neq x^*, (u(x)|(x - x^*)) > 0$.*

(A2) *$\sum_n \gamma_n = \infty$ and $\sum_n \gamma_n^2 < \infty$.*

(A3) *The function $x \mapsto \mathbb{E}(\|U(x, Z)\|^2)$ is bounded on any compact sets.*

Then, the sequence $(X_n)_n$ converges a.s. to x^ for any sequence of compact sets satisfying (3) and moreover the sequence $(\sigma_n)_n$ is a.s. finite (i.e. for n large enough $p_n = 0$ a.s.).*

We will not prove Theorem 1 directly as it actually derives from a more general result.

Theorem 2. *Under Hypothesis (A1) and if*

(A4) *$\sum_n \gamma_n = \infty$.*

(A5) *For all $q > 0$, the series $\sum_n \gamma_{n+1}\delta M_{n+1}\mathbb{1}_{\{\|X_n - x^*\| \leq q\}}$ converges almost surely.*

Then, the sequence $(X_n)_n$ converges a.s. to x^ and moreover the sequence $(\sigma_n)_n$ is a.s. finite (i.e. for n large enough $p_n = 0$ a.s.).*

Remark 4. *In the case where u derives from a potential V (i.e. $u = \nabla V$), Hypothesis (A1) is satisfied as soon as V is strictly convex.*

Proof of Theorem 1. It is sufficient to prove that the hypotheses of Theorem 1 imply the ones of Theorem 2. Consider $M_n = \sum_{i=1}^n \gamma_i \delta M_i \mathbb{1}_{\{\|X_{i-1} - x^*\| \leq q\}}$, $(M_n)_n$ is a martingale. By computing its angle bracket, we find $\langle M \rangle_n = \sum_{i=1}^n \gamma_i^2 \mathbb{E}(\delta M_i \delta M_i' | \mathcal{F}_{i-1}) \mathbb{1}_{\{\|X_{i-1} - x^*\| \leq q\}}$. As the series $\sum_i \gamma_i^2$ converges and the function $x \mapsto \mathbb{E}(\|U(x, Z)\|^2)$ is bounded on all compact sets, the almost sure convergence of $\langle M \rangle_n$ ensues from the Strong Law for square integrable martingales. Hence, we can apply Theorem 2, and the conclusion yields. \blacksquare

4. PROOF OF THEOREM 2

The proof of Theorem 2 is based on the following lemma which establishes a condition for the sequence $(X_n)_n$ to be a.s. compact.

Lemma 1. *If for all $q > 0$, the series $\sum_{n \geq 0} \gamma_n \delta M_n \mathbb{1}_{\{\|X_{n-1} - x^*\| < q\}}$ converges a.s. and if $p_n \mathbb{1}_{\{\|X_{n-1} - x^*\| < q\}} \rightarrow 0$, then the sequence $(X_n)_n$ remains a.s. in a compact set.*

Note that the compact set mentioned in Lemma 1 is random. In particular, this lemma does not imply that the number of truncations is bounded independently of the randomness ω .

Proof of Theorem 2. The proof is divided in two parts.

- Let $q > 0$. We define $\bar{M}_n = \sum_{i=1}^n \gamma_i \delta M_i \mathbb{1}_{\{\|X_{i-1} - x^*\| \leq q\}}$. Thanks to Hypothesis (A5), \bar{M}_n converges almost surely.

Assume that $\sigma_n \rightarrow \infty$. This is in contradiction with the conclusion of Lemma 1, which implies that the hypothesis according to which $p_n \mathbb{1}_{\{\|X_{n-1} - x^*\| < q\}}$ tends to 0 does not hold. So,

$$\exists \eta > 0, q > 0, \quad \forall N > 0, \exists n > N \quad \mathbb{1}_{\{\|X_n - x^*\| \leq q\}} \|p_{n+1}\| > \eta.$$

Let $\varepsilon > 0$. There exists a subsequence $X_{\phi(n)}$ such that for all $n > 0$, $\mathbb{1}_{\{\|X_{\phi(n)} - x^*\| \leq q\}} \|p_{\phi(n)+1}\| \neq 0$ and $\|\gamma_{\phi(n)+1} \delta M_{\phi(n)+1}\| \leq \varepsilon$.

So, $\|X_{\phi(n)} - x^*\| \leq q$ and however the new potential iterate $X_{\phi(n)+\frac{1}{2}} = X_{\phi(n)} - \gamma_{\phi(n)+1}(u(X_{\phi(n)}) + \delta M_{\phi(n)+1})$ is not in $\mathcal{K}_{\sigma_{\phi(n)}}$. Since u is continuous, $\|\gamma_{\phi(n)+1} u(X_{\phi(n)})\|$ can be made smaller than ε . As $\|\gamma_{\phi(n)+1} \delta M_{\phi(n)+1}\| \leq \varepsilon$, a proper choice of ε enables to write

$$\|X_{\phi(n)} - x^* - \gamma_{\phi(n)+1}(u(X_{\phi(n)}) + \delta M_{\phi(n)+1})\| \leq q + 1.$$

Let l be the smallest integer s.t. $B(x^*, q + 1) \subset \mathcal{K}_l$ (such an integer exists thanks to (3)), then $\sigma_{\phi(n)} < l$ for all n . Since the sequence $(\sigma_n)_n$ is increasing, this proves that $\limsup_n \sigma_n < \infty$ a.s..

- According to the previous item $\limsup_n \sigma_n < \infty$ a.s.. So, the sequence $(X_n)_n$ is almost surely compact. Consequently, we can in fact set $q = \infty$ in Hypothesis (A5) and say that $\sum_i \gamma_i \delta M_i$ converges almost surely. Let us consider

$$X'_n = X_n - \sum_{i=n+1}^{\infty} \gamma_i \delta M_i.$$

Since the series $\sum_{i \geq 0} \gamma_i \delta M_i$ converges a.s. and X_n remains in a compact set, X'_n also remains in a compact set. Let \mathcal{C} be this compact set. We define $\bar{u} = \sup_{x \in \mathcal{C}} \|u(x)\|$.

$$X'_{n+1} = X'_n - \gamma_{n+1} u(X'_n) + \gamma_{n+1} \varepsilon_n,$$

where $\varepsilon_n = u(X'_n) - u(X_n)$. Since $\|X'_n - X_n\| \rightarrow 0$ and u is continuous, $\|\varepsilon_n\| \rightarrow 0$.

$$\begin{aligned} \|X'_{n+1} - x^*\|^2 &\leq \|X'_n - x^*\|^2 - 2\gamma_{n+1}(X'_n - x^* | u(X'_n)) \\ &\quad + \gamma_{n+1}^2(\varepsilon_n^2 + \bar{u}^2) - 2\gamma_{n+1}(X'_n - x^* | \varepsilon_n). \end{aligned}$$

We can rewrite the inequality introducing a new sequence $\varepsilon'_n \rightarrow 0$.

$$\|X'_{n+1} - x^*\|^2 \leq \|X'_n - x^*\|^2 - 2\gamma_{n+1}(X'_n - x^* | u(X'_n)) + \gamma_{n+1} \varepsilon'_n. \quad (8)$$

Let $\delta > 0$. If $\|X'_n - x^*\|^2 > \delta$, then $(X'_n - x^*, u(X'_n)) > c > 0$. Henceforth, for n large enough Equation (8) becomes

$$\|X'_{n+1} - x^*\|^2 \leq \|X'_n - x^*\|^2 - \gamma_{n+1} c \mathbb{1}_{\{\|X'_n - x^*\|^2 > \delta\}} + \gamma_{n+1} (\bar{c} + \varepsilon'_n) \mathbb{1}_{\{\|X'_n - x^*\|^2 \leq \delta\}},$$

where $\bar{c} = \sup_{\|x - x^*\|^2 \leq \delta} (x - x^* | u(x))$. Since $\sum_n \gamma_n = \infty$, each time $\|X'_n - x^*\|^2 > \delta$, the sequence X'_n is driven back into the ball $\bar{B}(x^*, \sqrt{\delta})$ in a finite number of steps. Hence, for any n large enough

$$\|X'_n - x^*\|^2 < \delta + \gamma_{\phi(n)+1} (\bar{c} + \varepsilon'_{\phi(n)}),$$

where $\phi(n) = \sup\{p \leq n; \|X'_p - x^*\|^2 \leq \delta\}$. As $\phi(n)$ a.s. tends to infinity with n , $\limsup_n \|X'_n - x^*\|^2 \leq \delta$ for all $\delta > 0$. This proves that $X'_n \rightarrow x^*$. Finally, since the series $\sum_n \gamma_{n+1} \delta M_{n+1}$ converges, this also proves that $X_n \rightarrow x^*$. ■

Now, we are going to prove Lemma 1.

Proof of Lemma 1. If $\sigma_n < \infty$ a.s., the conclusion of the Lemma is obvious. Assume that $\sigma_n \rightarrow \infty$. Since each time σ_n increases, the sequence X_n is reset to a fixed point of \mathcal{K}_0 , the existence of a compact set in which the sequence lies infinitely often is straightforward.

Let $M > 0$, we set $\mathcal{C} = \{x; \|x - x^*\|^2 \leq M\}$. We can rewrite the Hypotheses of the Lemma as follows

$$\forall \varepsilon > 0, \exists N > 0 \text{ s.t. } \forall n, p \geq N \text{ we have } \begin{cases} \left\| \sum_{k=n}^p \gamma_k \delta M_k \mathbb{1}_{\{\|X_{k-1} - x^*\|^2 \leq M+2\}} \right\| < \varepsilon, \\ \gamma_n < \varepsilon, \\ \mathbb{1}_{\{\|X_{n-1} - x^*\|^2 \leq M+2\}} \|p_n\| < \varepsilon. \end{cases} \quad (9)$$

Let $\varepsilon > 0$ and $N > 0$ satisfying Condition (9) and s.t. $X_N \in \mathcal{C}$. We introduce

$$X'_n = X_n - \sum_{i=n+1}^{\infty} \gamma_i \delta M_i \mathbb{1}_{\{\|X_{i-1} - x^*\|^2 \leq M+2\}}.$$

By using Equation (5), we can easily show that X'_n satisfies the following recurrence relation

$$X'_{n+1} = X'_n - \gamma_{n+1} \delta M_{n+1} \mathbb{1}_{\{\|X_n - x^*\|^2 > M+2\}} - \gamma_{n+1} (u(X_n) - p_{n+1}). \quad (10)$$

We will now prove that the sequence $(X'_n)_n$ remains in the set $\{x; \|x - x^*\|^2 \leq M+1\} = \mathcal{C}'$.

The recurrent hypothesis is satisfied for $n = N$ (it is sufficient to choose $\varepsilon < 1$). Assume that the hypothesis holds for N, \dots, n . Hence, $\|X_n - x^*\|^2 \leq M+2$. Then, we can deduce from Equation (10) that

$$\begin{aligned} X'_{n+1} &= X'_n - \gamma_{n+1} (u(X_n) - p_{n+1}), \\ \|X'_{n+1} - x^*\|^2 &\leq \|X'_n - x^*\|^2 - 2\gamma_{n+1} (X'_n - x^* | u(X_n)) + \gamma_{n+1} c \varepsilon, \end{aligned}$$

where c is a positive constant independent of M .

- If $\|X'_n - x^*\|^2 \leq M$, thanks to the continuity of u , a proper choice of ε ensures that $\gamma_{n+1} \|(X'_n - x^* | u(X_n))\| < 1$. Hence, $\|X'_{n+1} - x^*\|^2 \leq M+1$.

- If $M < \|X'_n - x^*\|^2 \leq M + 1$, thanks to the continuity of u and thanks to Hypothesis (A1), $(X_n - x^* | u(X_n)) > \delta > 0$. Once again, properly choosing ε guaranties that $c\varepsilon < \delta$. Consequently, $\|X'_{n+1} - x^*\|^2 \leq M + 1$.

We have proved that for all $n > N$, $\|X'_n - x^*\|^2 \leq M + 1$. Since ε can be chosen smaller than 1, the following upper-bound also holds

$$\|X_n - x^*\|^2 \leq M + 2, \text{ for all } n > N.$$

This achieves to prove that the sequence $(X_n)_n$ remains in a compact set and consequently that $\limsup_n \sigma_n$ is a.s. finite. ■

Acknowledgement. I would like to thank Bernard Lapeyre for the fruitful remarks he made on a previous version of the proof presented above.

REFERENCES

- H. Chen and Y. Zhu. *Stochastic Approximation Procedure with randomly varying truncations*. Scientia Sinica Series, 1986.
- H. F. Chen, G. Lei, and A. J. Gao. Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.*, 27(2):217–231, 1988. ISSN 0304-4149.
- B. Delyon. General results on the convergence of stochastic algorithms. *IEEE Transactions on Automatic Control*, 41(9):1245–1255, 1996.
- B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.*, 27(1):94–128, 1999. ISSN 0090-5364.
- H. J. Kushner and D. S. Clark. *Stochastic approximation methods for constrained and unconstrained systems*, volume 26 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1978. ISBN 0-387-90341-0.
- H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statistics*, 22:400–407, 1951.

CERMICS, ECOLE DES PONTS, PARISTECH, 6-8 AVENUE BLAISE PASCAL, CHAMPS SUR MARNE 77455 MARNE LA VALLÉE, FRANCE.

E-mail address: lelong@cermics.enpc.fr